Cloud computing for Astroscience applications

Cécile Cavet, Michèle Detournay, Volker Beckmann

Cloud computing for Astroscience applications

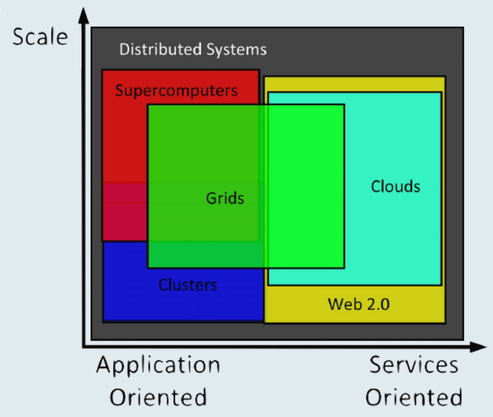
Cécile Cavet, Michèle Detournay, Volker Beckmann
François Arago Centre, APC, Univ. Paris Diderot, CNRS/IN2P3, CEA/Irfu, Obs. de Paris, Sorbonne Paris Cité,
13 rue Watt, 75013, Paris, France
cecile.cavet@apc.univ-paris7.fr

Abstract

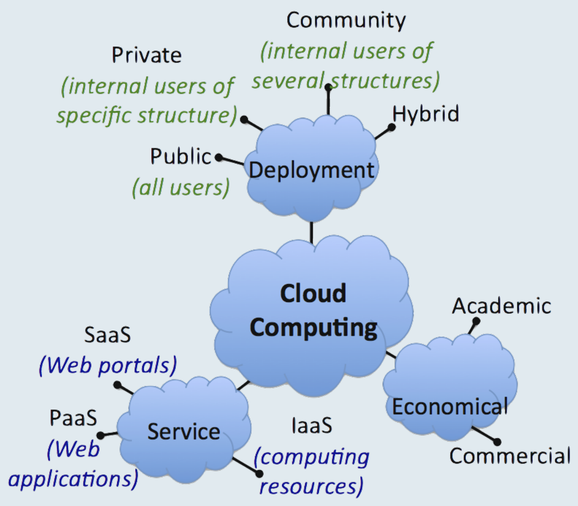
Cloud computing offers virtual resources on-demand. The Infrastructure-as-a-Service (IaaS) Cloud allows to create virtual machines (VMs), virtual storage and virtual network. We present here main properties and instructions for the use of Cloud computing in the framework of the academic StratusLab Cloud. In the scientific domain of Astroscience applications, we have conducted benchmark performance tests to mesure the overhead due to virtualisation. We also performed related studies such as multi-cloud solution and cluster for Big Data.

Definition

Cloud computing is a new IT paradigm taking place in the distributed computing family.

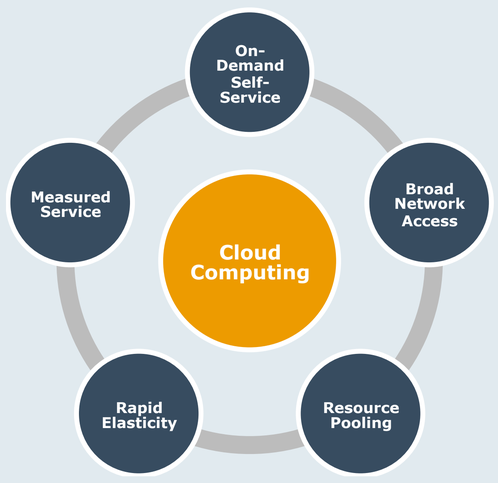


Distributed computing [4]: the Cloud computing paradigm has emerged due to Web evolution and virtualisation.



Cloud computing models.

The main attribute of Cloud computing is its flexibility.

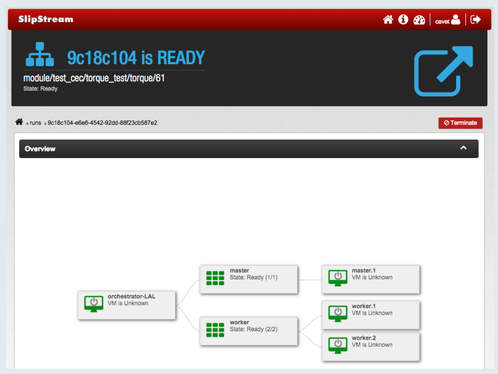


► Due to all these properties, Cloud computing is a promising platform for **scientific computing**.

SlipStream: multi-cloud

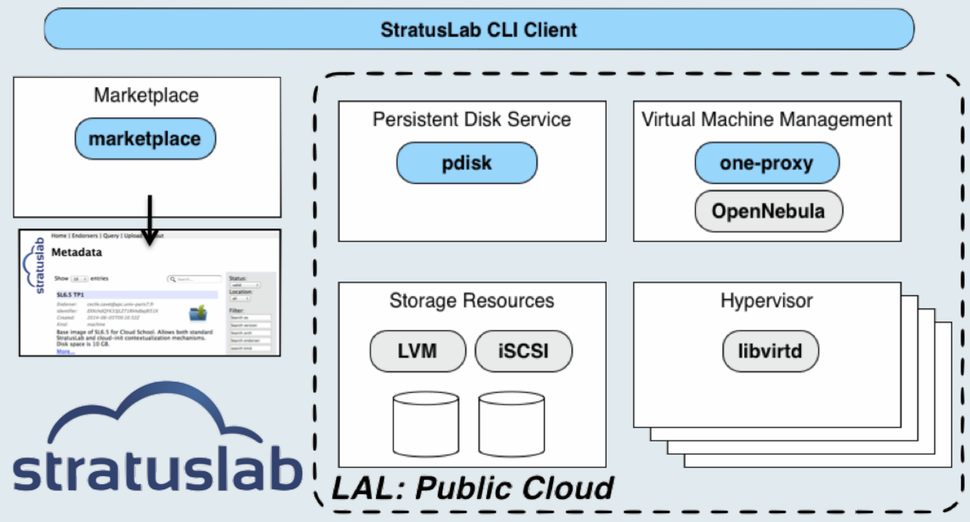
The **SlipStream™** solution [9] allows to automatically deploy an environnement on VMs. With the Web interface, users just need to:

- create an SlipStream account and register Cloud credentials and ssh keys.
- run VM or deployment and access by ssh to VM.



SlipStream Web interface: example of Torque deployment on a virtual cluster of one master node and two worker nodes.

StratusLab Cloud



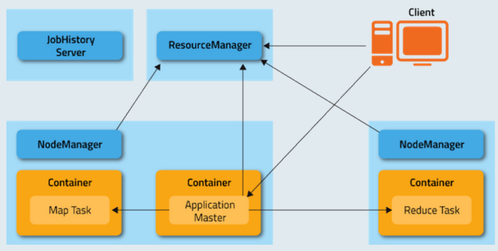
Working of the StratusLab Cloud [8].

The **StratusLab** [10] project offers an academic IaaS Cloud since 2010. This public Cloud provides both **StratusLab** Cloud solution and virtual computing resources. VMs are:

- generated by **KVM** hypervisor and managed by **OpenNebula** virtual infrastructure manager (sonly replaced by **LibVirt**).
- supported by physical machines @LAL: **16 nodes**, **440 cores, 772 GB of memory** and **15 TB of storage**.
- provided in OS disk images by the **MarketPlace** [7] and manageable by the **StratusLab command line client**.

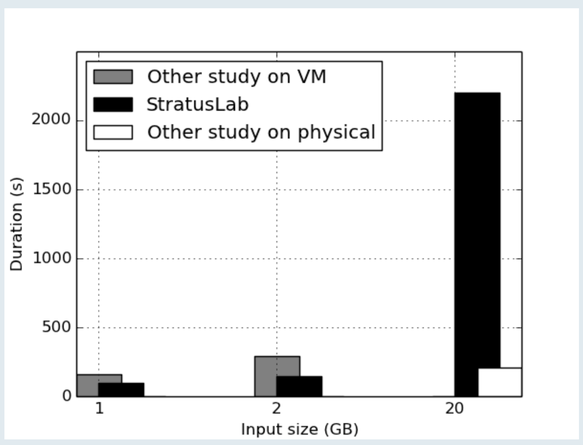
Big Data on the Cloud

Technics to treat huge volume of data are encompassed in **Big Data** term which includes algorithme (**Map/Reduce**) and environment (**Hadoop cluster**). Hadoop cluster is a challenging application due to huge amount of memory and storage requirements.



Hadoop Cluster with YARN (2.0)

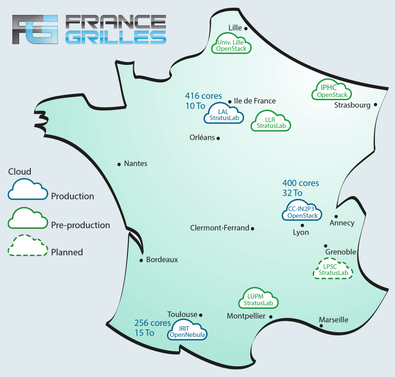
Cluster benchmarking with Sort application is the first step of the study.



Sort benchmark (preliminary result).

French Cloud Federation

The French actors of academic Cloud computing are federated [1] in order to offer Cloud services to user. Several sites provide **StratusLab**, **OpenStack** and **OpenNebula** with specific resources.



French Cloud infrastructures [3].

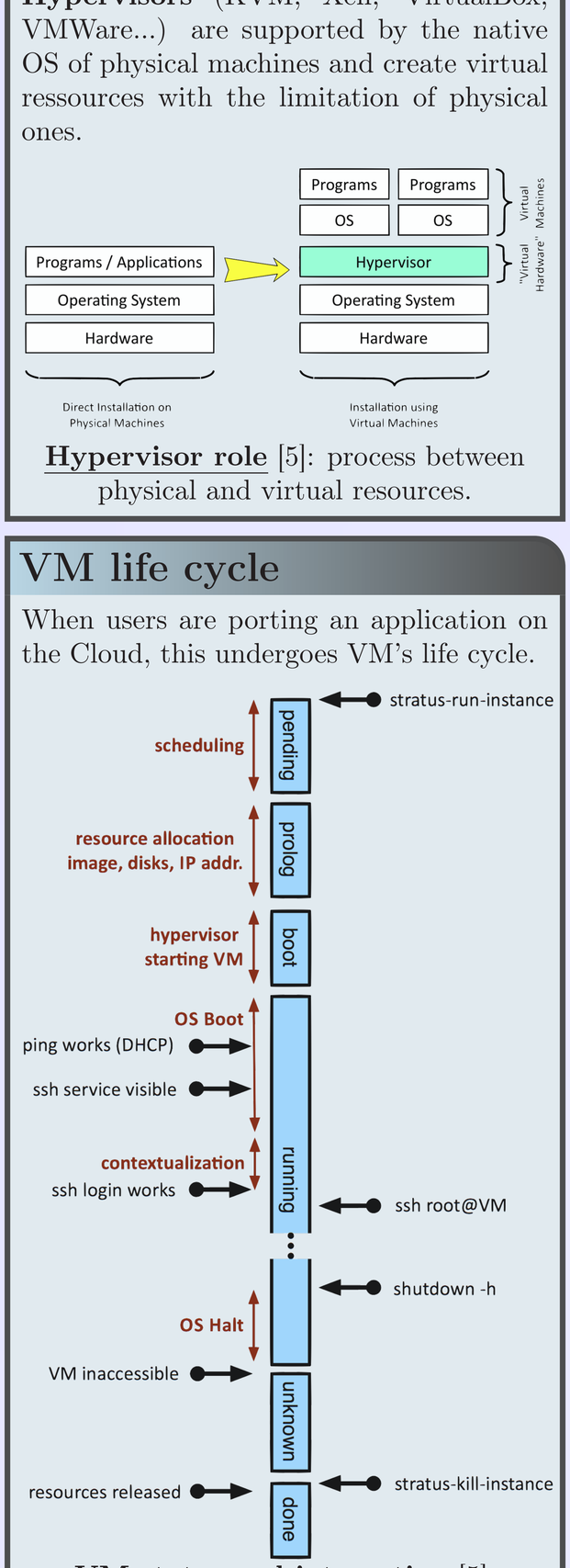
Virtualisation

Virtualisation is at the heart of the process allowing for the creation of VMs. **Hypervisors** (KVM, Xen, VirtualBox, VMWare...) are supported by the native OS of physical machines and create virtual ressources with the limitation of physical ones.

Hypervisor role [5]: process between physical and virtual resources.

VM life cycle

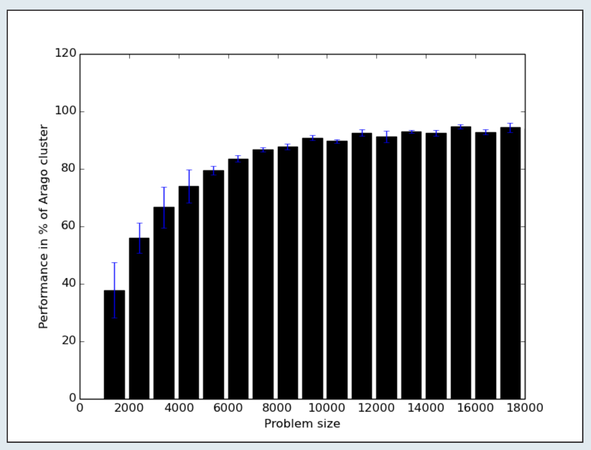
When users are porting an application on the Cloud, this undergoes VM's life cycle.

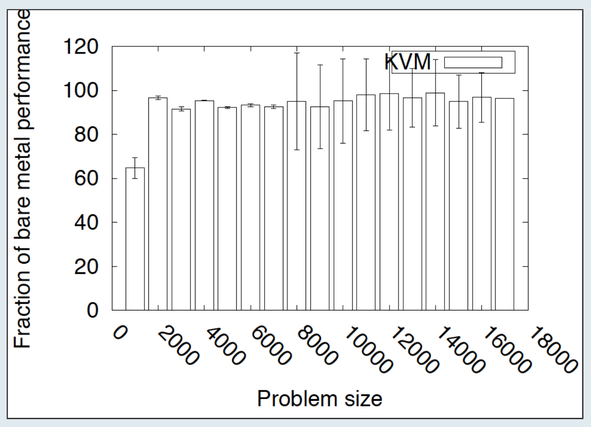


VM states and interaction [5].

Performance Benchmark

To obtain VM performance, **High Performance Computing Challenge** (HPCC) benchmark was run [2] and particularly the **High Performance LINPACK** (HPL) benchmark. Small differences are due to the type of CPU (cache size).





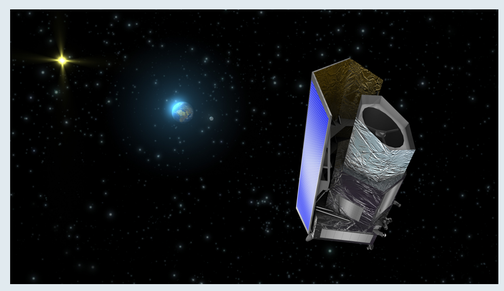
HPL benchmark: performance comparaison of the Arago cluster (FACe/APC) vs StratusLab VM (top) and a bare metal system vs KVM VM (bottom) [6].

► The behaviour is similar: no **overhead** due to virtualisation.

References

[1] Airaj et al., hal-00927506 (2013)
[2] Cavet et al., hal-00766067 (2012)
[3] Cloud Federation: www.france-grilles.fr/6-Cloud
[4] Foster et al., IEEE (2009)
[5] Loomis, Journée Cloud (2012)
[6] Luszczek et al. (2011)
[7] Marketplace: marketplace.stratuslab.eu/marketplace/metadata
[8] Philippon, Ecole informatique IN2P3 (2014)
[9] SlipStream: sixsq.com/products/slipstream.html
[10] StratusLab: stratuslab.eu

Application porting



Euclid mission (@Euclid).

► **Application porting = adaptation of the IT environnement to the code**.

► **Cloud computing advantages**:

- OS and resources on-demand.
- Immediate instantiation and ssh connection.
- Infinite life duration of the VM (or almost).
- Advanced system using due to root user rights.
- Immediate code porting and execution.
- No overhead on computing (execution time is similar).